The Distributions of Freeze-date and Freeze-free Period When a Probability of No Freeze Exists

by H.C.S. Thom *

The estimation of freeze probabilities from complete freeze-date series has been treated by Thom and Shaw.[1] When the freeze-date series for an observation station is incomplete in the sense that some years experienced no freeze, a probability of no freeze occurring at all exists. This often occurs in more southerly latitudes especially with freeze thresholds below 32° F. With the addition of the no-freeze probability component a quite different problem in the estimation of freeze probability arises. This has been discussed by Spillman _et_. _al_.[2] They gave rules for finding the mean recurrence interval for the incomplete series but since they did not recognize the more general statistical aspects of the problem their rules are not completely convertible to probability statements.

The model for determining freeze probability may be thought of as a mixture of two distributions: one a discrete distribution of no-freeze and freeze, the other a continuous distribution of freeze-date when freeze occurred. The period over which spring freeze-date is assumed to range is January 1 to June 30 and for fall freeze from July 1 to December 31. These are arbitrary and other dates may be assumed if it suits a particular purpose better. The model is seen to be equivalent to concentrating a probability of no freeze at an arbitrary point before the beginning of the freeze season for spring freeze and after the freeze season for fall freeze.

---

* Visiting Professor of Statistics, Biometrics Unit, Cornell University; Chief Climatologist, U.S. Weather Bureau.

Let q_s be the probability of no spring freeze, p_s the probability of spring freeze, and $F_s(x)$ the distribution function of spring freeze-date, i.e., the probability of freeze before date x given that freeze will occur. Then the probability of freeze before a date x in spring is

$$G_s(x) = q_s + p_s F_s(x) \tag{1}$$

hence the probability of freeze after a date x is

$$H(x) = 1 - G_s(x) = 1 - q_s - p_s F_s(x)$$

Since $p_s + q_s = 1$ this may be written

$$H(x) = p_s \left[1 - F_s(x)\right] \tag{2}$$

For fall-freeze we have the mixed distribution function

$$H(x) = q_a + F_a(x) \tag{3}$$

and for the probability of freeze after date x

$$G_a(x) = 1 - \left\{q_a + p_a \left[1 - F_a(x)\right]\right\} \tag{4}$$

where q_a is the probability of no fall freeze and p_a the probability of fall freeze.

This reduces to

$$G_a(x) = p_a F_a(x) \tag{5}$$

a form analogous to (2). It will be noted that when p_s and p_a are unity (2) and (4) reduce to the small equations for freeze probability.

p may be estimated by the equation

$$\hat{p} = \frac{m}{n} \tag{6}$$

where m is the number of years with freeze and n the total number of years. F(x) may be estimated without defining F(x) analytically by using the equations

$$\hat{F}(x) = \frac{r}{m} \tag{7}$$

where r is the number of the freeze in a series ordered by date. It may also be estimated by assuming some analytical form for F(x). Both procedures will be illustrated on records for Anniston, Auburn, and Birmingham, Alabama using the 16°, 24°, and 20° thresholds respectively.

Freeze Probabilities, no Date Distribution Assumed

Tables 1, 2, and 3 give the freeze series coded to March 1 arranged in order of date together with estimated probabilities F(x), G(x), and H(x). F(x) is estimated by equation (7), H(x) by equation (2), and G(x) by equation (5). The results are shown as distribution functions or ogives in Figures 1, 2, and 3. To obtain the probability of freeze after any date in spring and before any date in fall, one merely finds the date on the abscissa and reads off the probability on the ordinate.

Freeze Probabilities, Date Distribution Assumed Normal

W.G. Reed[3] and later Thom and Shaw[1] found that the normal distribution adequately fitted freeze-date series under a wide range of conditions for the 32° threshold. Since the effect of the null component of probability in the mixed distribution might be a truncation of the freeze date series on the left in spring and on the right in fall, it seemed desirable to test for departure from normality. Methods provided by Geary and Pearson[4] were used to make these tests. All values of a and $\sqrt{b_1}$ of table 4 are either in the acceptance region given by their tables or, where the sample size is smaller than tabled, they are in the shorter acceptance regions for the smallest sample size tabulated, with one exception. This is the skewness for 28° fall freeze at Anniston. Since other freeze thresholds exhibited less skewness, particularly the lower thresholds where truncation would have a more marked effect, the one significant value was not deemed to be sufficient evidence against normality.

Table 1

Estimated Probabilities

Anniston 16°

Spring

r	Date	1-F(x)	1-n(x)	|d|	H(x)	N(x)
1	-52	.937	.934	.003	.517	.516
2	-48	.875	.913	.038	.483	.504
3	-43	.812	.873	.061	.448	.482
4	-34	.750	.770	.020	.414	.425
5	-31	.687	.729	.042	.379	.402
6	-28	.625	.681	.056	.345	.376
7	-18	.562	.512	.050	.310	.283
8	-18	.500	.512	.012	.276	.283
9	-17	.437	.492	.055	.241	.272
10	-15	.375	.460	.085	.207	.254
11	-9	.312	.366	.054	.172	.202
12	-9	.250	.366	.116	.138	.202
13	-1	.177	.233	.056	.098	.129
14	0	.125	.221	.096	.069	.122
15	10	.062	.111	.059	.034	.061
16	35	.000	.010	.010	.000	.005

Fall

r	Date	F(x)	n(x)	|d|	G(x)	N(x)
1	260	.111	.018	.093	.036	.006
2	277	.222	.251	.029	.071	.081
3	281	.333	.367	.034	.107	.118
4	283	.444	.433	.011	.143	.139
5	286	.556	.532	.024	.178	.171
6	290	.667	.663	.004	.214	.213
7	291	.778	.691	.087	.250	.222
8	297	.889	.844	.045	.285	.271
9	300	1.000	.896	.104	.321	.288

Table 2

Estimated Probabilities

Birmingham 20°

Spring

r	Date	1-F(x)	1-n(x)	|d|	H(x)	N(x)
1	-48	.947	.970	.023	.599	.614
2	-44	.895	.952	.057	.567	.603
3	-35	.842	.875	.033	.533	.554
4	-31	.789	.821	.032	.499	.520
5	-29	.737	.791	.054	.467	.501
6	-28	.684	.773	.089	.433	.489
7	-18	.632	.571	.061	.400	.361
8	-17	.579	.548	.031	.367	.347
9	-15	.526	.504	.022	.333	.319
10	-13	.474	.456	.018	.300	.289
11	-9	.421	.367	.054	.266	.232
12	-9	.368	.367	.001	.233	.232
13	-9	.316	.367	.051	.200	.232
14	-9	.263	.367	.104	.166	.232
15	0	.211	.198	.013	.134	.125
16	3	.158	.154	.004	.100	.097
17	4	.105	.140	.035	.066	.089
18	10	.053	.076	.028	.034	.048
19	14	.000	.049	.018	.000	.031

Fall

r	Date	F(x)	n(x)	|d|	G(x)	N(x)
1	259	.083	.021	.062	.033	.008
2	268	.167	.109	.058	.067	.044
3	273	.250	.215	.035	.100	.086
4	276	.333	.302	.031	.133	.121
5	280	.417	.433	.016	.167	.173
6	282	.500	.504	.004	.200	.202
7	285	.583	.606	.023	.233	.242
8	289	.667	.736	.069	.267	.294
9	290	.750	.764	.014	.300	.306
10	290	.833	.764	.069	.333	.306
11	295	.917	.877	.040	.367	.351
12	296	1.000	.894	.106	.400	.358

Table 3

Estimated Probabilities

Auburn 24°

Spring

r	Date	1-F(x)	1-n(x)	|d|	H(x)	N(x)
1	-58	.963	.980	.017	.867	.882
2	-57	.926	.977	.051	.833	.879
3	-44	.889	.918	.029	.800	.826
4	-42	.852	.903	.051	.767	.813
5	-35	.815	.834	.019	.734	.751
6	-31	.778	.785	.007	.700	.707
7	-27	.741	.726	.015	.667	.653
8	-27	.704	.726	.022	.634	.653
9	-25	.667	.695	.028	.600	.626
10	-20	.630	.606	.024	.567	.545
11	-19	.593	.591	.002	.534	.532
12	-17	.556	.552	.004	.500	.497
13	-16	.519	.536	.017	.467	.482
14	-15	.481	.516	.035	.433	.464
15	-9	.444	.405	.039	.400	.365
16	-9	.407	.405	.002	.366	.365
17	-5	.370	.324	.036	.333	.301
18	0	.333	.255	.078	.300	.230
19	0	.296	.255	.041	.266	.230
20	1	.259	.239	.020	.233	.215
21	3	.222	.212	.010	.200	.191
22	4	.185	.198	.013	.167	.178
23	8	.148	.149	.001	.133	.134
24	11	.111	.119	.008	.100	.107
25	13	.074	.102	.028	.067	.092
26	14	.037	.093	.056	.033	.084
27	20	.000	.055	.055	.000	.050

Fall

r	Date	F(x)	n(x)	|d|	G(x)	N(x)
1	259	.053	.077	.024	.034	.048
2	263	.105	.117	.012	.066	.074
3	264	.158	.133	.025	.100	.084
4	268	.211	.212	.001	.134	.134
5	269	.263	.233	.030	.166	.147
6	270	.316	.258	.058	.200	.163
7	271	.368	.284	.084	.233	.180
8	272	.421	.312	.109	.266	.197
9	273	.474	.337	.137	.300	.213
10	274	.526	.367	.159	.333	.232
11	276	.579	.429	.150	.367	.272
12	283	.632	.641	.009	.400	.406
13	283	.684	.641	.043	.433	.406
14	285	.737	.695	.042	.467	.440
15	288	.789	.773	.016	.499	.489
16	290	.842	.816	.026	.533	.517
17	297	.895	.925	.030	.567	.586
18	300	.947	.954	.007	.599	.604
19	304	1.000	.977	.023	.633	.618

Tablo 4

Anniston

Spring n = 29

Freeze	m	$\bar{x}$ code	$\bar{x}$ date	S	p	a	$\sqrt{b_1}$
32	29	30.1	3/30	13.7	1.000	.820	-.295
28	29	9.8	3/10	16.0	1.000	.768	.005
24	28	-4.4	2/24	18.1	.966	.779	-.846
20	26	-12.0	2/16	19.1	.897	.816	-.375
16	16	-17.4	2/11	22.5	.552	.736	.431

Fall n = 28

Freeze	m	$\bar{x}$ code	$\bar{x}$ date	S	p	a	$\sqrt{b_1}$
32	28	250.0	11/6	12.0	1.000	.795	.310
28	28	260.5	11/17	11.4	1.000	.770	-.985
24	22	273.1	11/29	13.9	.786	.716	.888
20	14	277.8	12/4	10.9	.500	.829	.279
16	9	285.0	12/11	11.9	.321	.728	-.710

Auburn

Spring n = 30

Freeze	m	$\bar{x}$ code	$\bar{x}$ date	S	p	a	$\sqrt{b_1}$
32	30	21.1	3/21	16.4	1.000	.805	-.219
28	30	.4	2/28	18.0	1.000	.793	-.478
24	27	-14.2	2/14	21.4	.900	.823	-.381
20	23	-16.5	2/11	19.5	.767	.864	-.173
16	12	-24.4	2/4	15.3	.400	.789	.244

Fall n = 30

Freeze	m	$\bar{x}$ code	$\bar{x}$ date	S	p	a	$\sqrt{b_1}$
32	29	257.7	11/14	9.0	.967	.842	-.203
28	27	272.9	11/29	11.4	.900	.725	.910
24	19	278.4	12/4	12.9	.633	.841	.468
20	10	280.0	12/6	11.4	.333	.807	-.194
16	6	287.0	12/13	11.3	.200	.738	-.266

Birmingham

Spring n = 30

Freeze	m	$\bar{x}$ code	$\bar{x}$ date	S	p	a	$\sqrt{b_1}$
32	30	18.4	3/18	16.1	1.000	.764	-.493
28	30	3.6	3/4	17.1	1.000	.770	.097
24	28	-7.7	2/20	17.2	.933	.818	-.245
20	19	-14.9	2/14	17.5	.633	.788	-.219
16	14	-24.8	2/3	19.6	.267	.849	.170

Fall n = 30

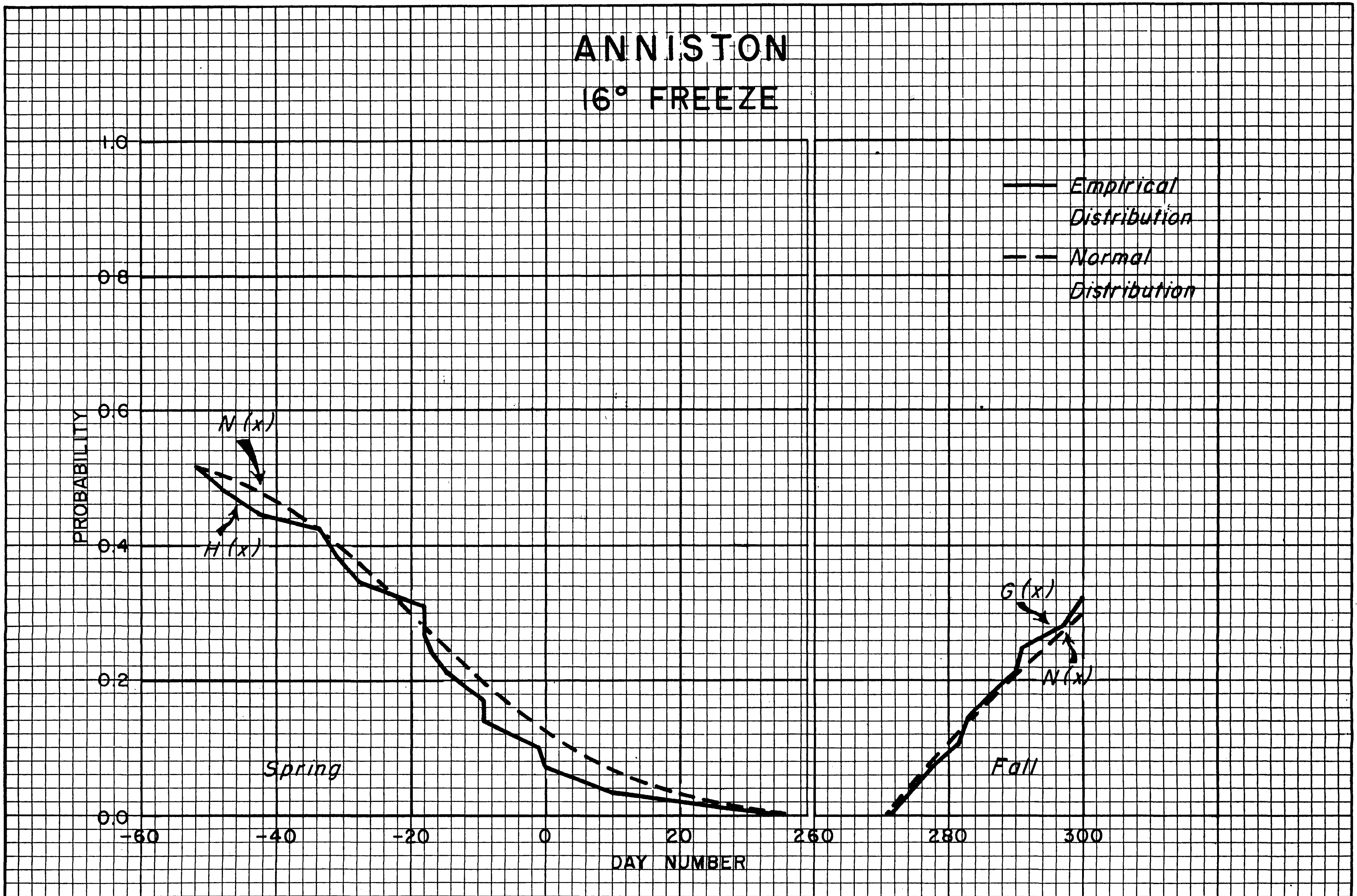
Freeze	m	$\bar{x}$ code	$\bar{x}$ date	S	p	a	$\sqrt{b_1}$
32	29	256.4	11/12	12.0	.967	.797	-.019
28	28	274.5	12/1	14.2	.933	.825	.116
24	21	279.4	12/5	12.5	.700	.854	.362
20	12	281.9	12/8	11.3	.400	.790	-.531
16	8	286.5	12/13	11.6	.267	.819	-.135

Normal distributions n(x) were fitted to the freeze-date series in the usual manner[1] by estimating means and standard deviations. The results are given in Tables 1, 2, and 3. N(x) is the probability of freeze after the corresponding date in spring and before the corresponding date in fall obtained from equations (2) and (4) on the assumption that $F_s(x)$ and $F_a(x)$ are normal. The spring and fall N(x) columns correspond to the H(x) and G(x) columns where the probabilities were estimated non-parametrically. The normal mixed distributions are also shown in Figures 1, 2, and 3.

Although the fit of the normal distributions appears to be satisfactory, it seemed desirable to make a further test of the agreement between the normal and non-parametric fits. For this purpose we use the easily applied Kolmogorov-Smirnov test for which Massey[5] has provided a very convenient set of tables. To apply this test it is only necessary to determine the maximum absolute difference in probability between the normal and empirical distributions. The absolute differences for the distributions of the date component $|d| = F(x) - n(x)$ are given in Tables 1, 2, and 3. The maximum value $|d|$ for each of the tables could be tested separately, but since the fits are good, it suffices to test the maximum for all tables. This value is seen to be 0.159 and occurs in fall 24°-distribution for Auburn. Referring to Massey's table for n = 30, a sample size greater than any encountered in the three tables and therefore a more stringent test, we find

$$P\left[\max |d| > .159\right] > 0.20.$$

It may be concluded therefore that the differences between normal and empirical fits could easily have resulted from random sampling and that the normal distribution provides a good fit to the freeze series. This goodness of fit is extended to the mixed distributions N(x) since there is no question as to the fit of the p's.



ANNISTON
16° FREEZE
Empirical Distribution
Normal Distribution
PROBABILITY
1.0
0.8
0.6
0.4
0.2
0.0
N (x)
H (x)
G (x)
N (x)
Spring
Fall
-60
-40
-20
0
20
260
280
300
DAY NUMBER

FIG. I

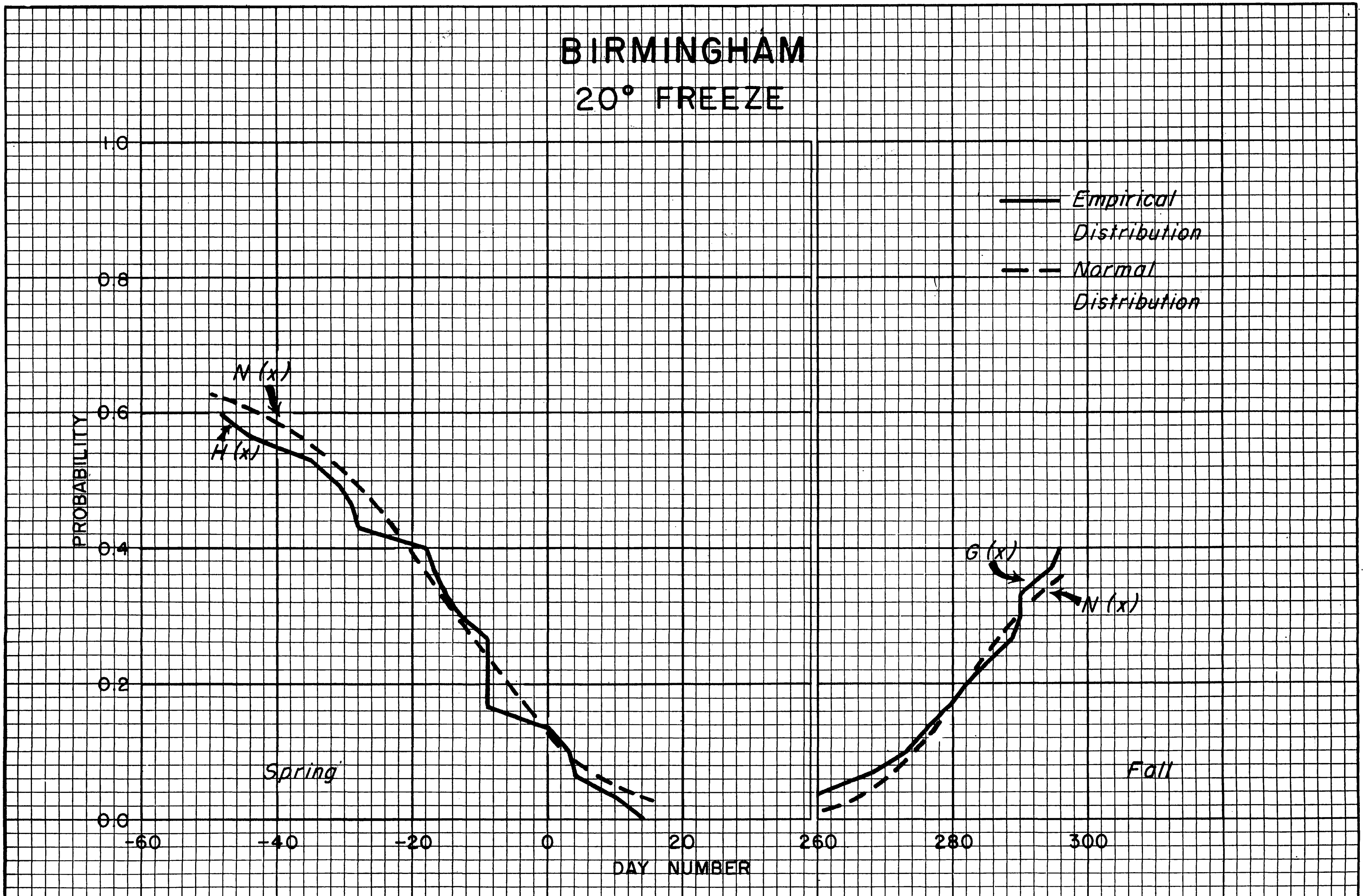

BIRMINGHAM
20° FREEZE
Empirical Distribution
Normal Distribution
N (x)
H (x)
G (x)
N (x)
Spring
Fall
PROBABILITY
1.0
0.8
0.6
0.4
0.2
0.0
-60
-40
-20
0
20
260
280
300
DAY NUMBER

FIG. 2

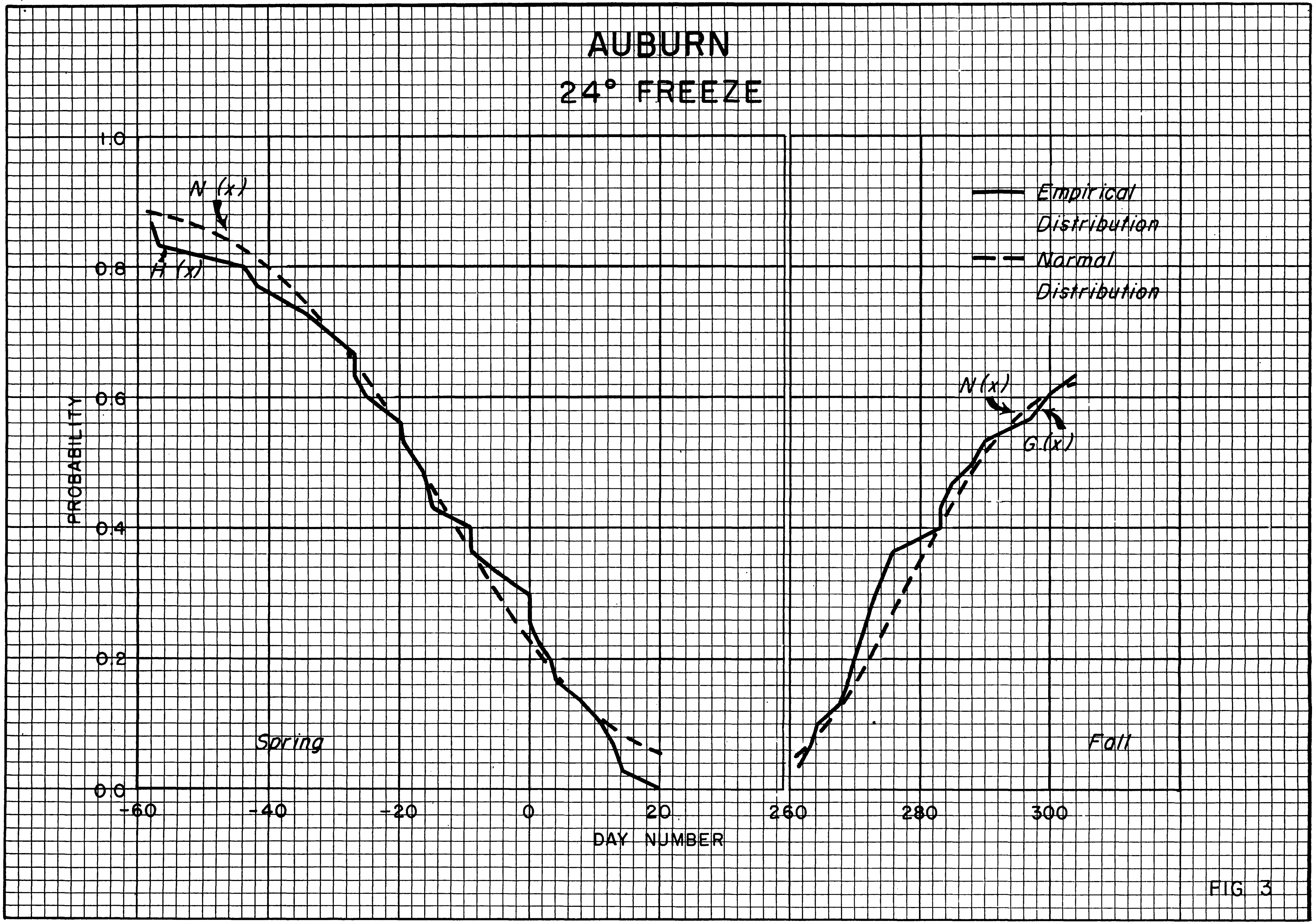

AUBURN
24° FREEZE
1.0
0.8
0.6
0.4
0.2
0.0
PROBABILITY
N (x)
H (x)
G (x)
Empirical Distribution
Normal Distribution
Spring
Fall
-60
-40
-20
0
20
260
280
300
DAY NUMBER

FIG 3

Freeze-free Season Probabilities

Since the freeze-free period is $y = x_a - x_s$, it is clear that the distribution of the freeze-free period will be a composite of the spring and fall distributions. We shall assume based on previous work that spring and fall freeze are distributed independently and we will find the distribution of $x_a - x_s$, the freeze-free period. To accomplish this it seems simplest to use characteristic functions.[6] The characteristic functions of (1) and (3) are

$$f(u) = q_s + p_s f_s(t) \tag{8}$$

and

$$f(v) = q_a + p_a f_a(t) \tag{9}$$

where f is a characteristic function and u, v, and t are arbitrary variables of the Fourier integrals. Since the characteristic function of the difference of two independent variables is the product of their characteristic functions we have

$$f(w) = q_s q_a + q_a p_s f(t_s) + q_s p_a f(t_a) + p_s p_a f_s(t) f_a(t) \tag{10}$$

the characteristic function of the freeze-free distribution. It is well known that the characteristic function completely determines its distribution hence

$$J(y) = q_s q_a + q_a p_s F_s(x) + q_s p_a F_a(x) + p_s p_a F_{sa}(y)$$

is the distribution of freeze-free period. In order to give this equation a reasonable interpretation, it must be agreed that if no freeze occurs in spring $y = x_a$, if no freeze occurs in fall $y = 365 - x_s$. Thus the freeze-free period is limited to 365 days; if no freeze occurs in spring it will begin on January 1 and if no freeze occurs in fall it will end on December 31. Other dates could be assumed if it seemed appropriate. It should be noted that the x employed here is the day number starting from January 1.

Making the substitutions of y for x we find

$$J(y) = q_s q_a + q_a p_s F_s(y) + q_s p_a F_a(y) + p_s p_a F_{sa}(y) \qquad (11)$$

the distribution of freeze-free period as a function of y.

In order to evaluate (11) it is necessary to assume the form of F_s and F_a. Since the y's are linear transformations of the x's, if we assume normality for $F_s(x)$ and $F_a(x)$, $F_s(y)$ and $F_a(y)$ will also be normal with the same standard deviations and means $\bar{y}_s$ and $\bar{y}_a$. $F_{sa}(y)$ will be a normal distribution with mean $\bar{x}_a - \bar{x}_b$ and standard deviation $\sqrt{s_s^2 + s_a^2}$. The q's and p's are defined as previously and the distribution is again a mixture of freeze and no-freeze components. The first term is the probability of no-freeze or a freeze-free period of 365 days; the second term the probability for a period January 1 to x_a; the third for a period x_s to December 31; and the last for a period with freeze at both ends.

Table 5 gives the mixture components and the freeze-free distribution J(y). $F_s(y)$ and $F_a(y)$ were obtained from the n(x) of Table 1 by transformation to the freeze-free period scale. These were weighted by the qp products and added to obtain the mixed distribution. J(y) is shown in Figure 4 from which the probability of the 16°-freeze-free period being less than any value y may be obtained. Freeze-free periods for other threshold values may also be readily computed including those with different thresholds in spring and fall.

Since (11) is a distribution function the expected value of y or the mean freeze-free period may be evaluated. This is found to be

$$E(y) = 360\, q_s q_a + q_a p_s E(y_s) + 2_s p_a E(y_a) + p_s p_a E(y_{sa}) .$$

Substituting the estimates for the expected values we find

$$\bar{y} = 360\, q_s q_a + q_a p_s \bar{y}_s + q_s p_a \bar{y}_a + p_s p_a \bar{y}_{sa} . \qquad (12)$$

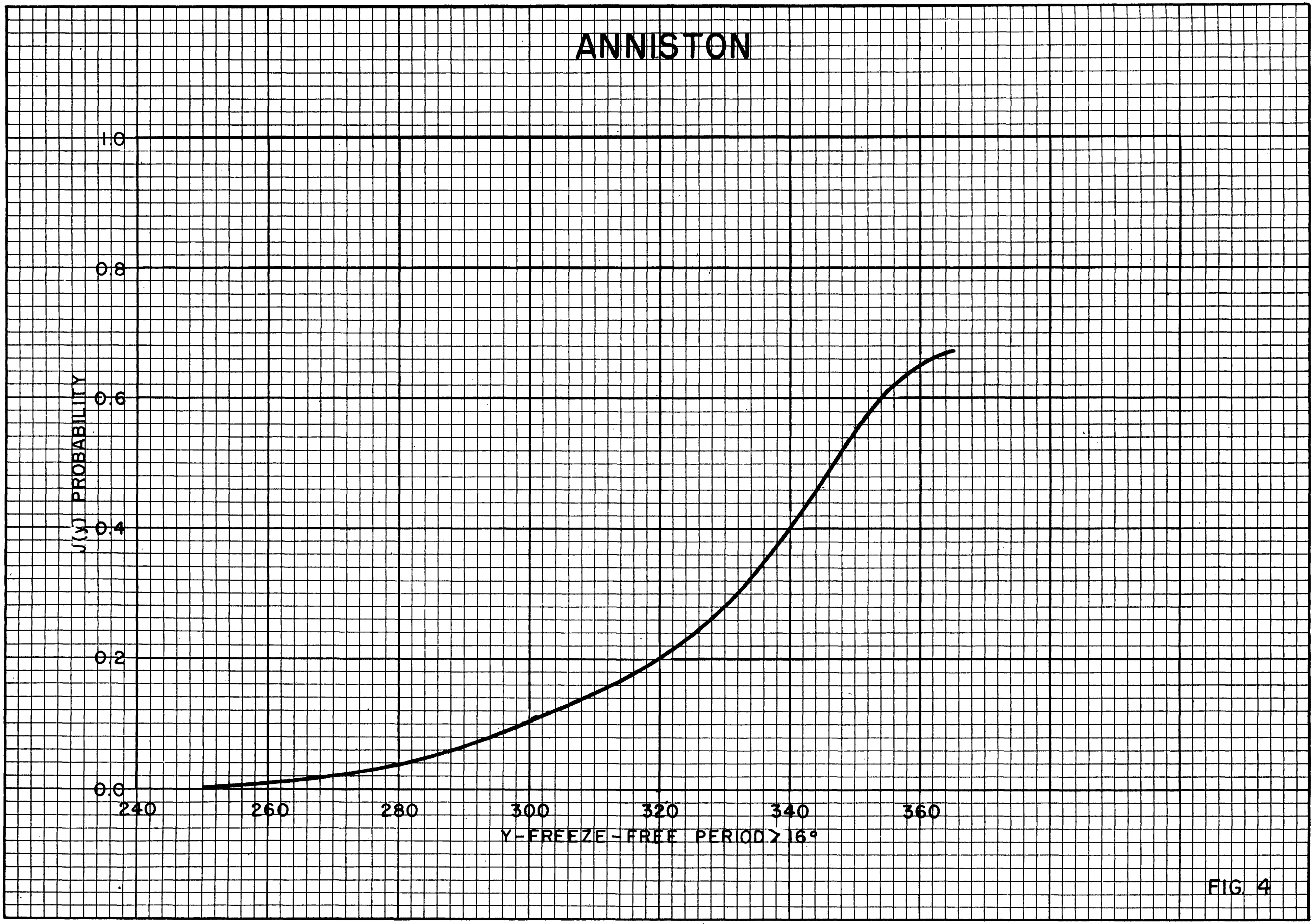

ANNISTON
J(y) PROBABILITY
1.0
0.8
0.6
0.4
0.2
0.0
240
260
280
300
320
340
360
Y-FREEZE-FREE PERIOD > 16°
FIG. 4

Table 5

16°-Freeze-free Distribution for Anniston

y	$F_{sa}(y)$	$F_s(y)$	$F_a(y)$	$p_a p_a F_{sa}$	$q_a p_s F_s$	$q_s p_a F_a$	$J(y)$
250	.020	.001		.004			.004
255	.031	.001		.006			.006
260	.047	.002		.008			.008
265	.071	.005		.013	.000		.013
270	.100	.009		.018	.001		.019
275	.140	.016		.025	.002		.027
280	.189	.027		.033	.004		.037
285	.245	.044		.043	.006		.047
290	.312	.069		.055	.010		.065
295	.386	.104		.068	.015		.083
300	.464	.149	.000	.082	.021		.103
305	.540	.204	.001	.096	.029		.125
310	.618	.274	.002	.109	.039	.001	.149
315	.691	.356	.007	.122	.051	.003	.176
320	.755	.440	.022	.134	.063	.008	.205
325	.813	.470	.055	.144	.068	.021	.233
330	.862	.614	.119	.153	.088	.045	.286
335	.900	.698	.224	.159	.101	.084	.344
340	.931	.770	.367	.165	.111	.138	.414
345	.954	.831	.468	.169	.120	.176	.465
350	.969	.881	.691	.172	.127	.259	.558
355	.981	.919	.821	.174	.132	.308	.614
360	.988	.948	.910	.175	.137	.341	.653
365	.993	.969	.961	.176	.140	.360	.676
366							1.000

$\bar{y}_s = 323.4$ $\qquad q_s q_a = .304$

$\bar{y}_a = 344.0$ $\qquad q_a p_s = .144$

$\bar{y}_{sa} = 302.4$ $\qquad q_s p_a = .375$

$s_{sa} = 25.4$ $\qquad p_s p_a = .177$

For Anniston 16°-freeze we readily find, using the values in Table 5 in (12), y = 340.

Summary

It has been shown that the freeze distribution is a mixture of the distributions of freeze-date and freeze-no freeze events. This has been applied both non-parametrically and assuming normal distribution of freeze date to three stations at three different thresholds to obtain the probabilities of freeze before or after any date. The distribution of the freeze-free period has been found and application made to one of the stations to obtain probabilities of the freeze-free period being less than a given length. The expression for the mean freeze-free period was also found and an estimate made of its value for the same station.

References

1. Thom, H.C.S. and Shaw, R.H. "Evaluation of Spring and Fall Freeze Hazard." (Manuscript)

2. Spillman, W.J., Tolley, H.R., and Reed, W.G. "The Average Interval Curve and its Application to Meteorological Phenomena." Monthly Weather Review, vol. 44, 1916, p. 197.

3. Reed, W.G. "The Problem Growing Season." Monthly Weather Review, vol. 44, p. 509.

4. Geary, R.C. and Pearson, E.S. "Tests of Normality." Biometrika Office, London (1938).

5. Massey, F.J.,Jr. "The Kolmogorov-Smirnov Test for Goodness of Fit." Jour. Am. Stat. Assoc., v. 46, (March 1951), p. 68.

6. Cramér, Harald. "Mathematical Methods of Statistics." Princeton University Press (1946).

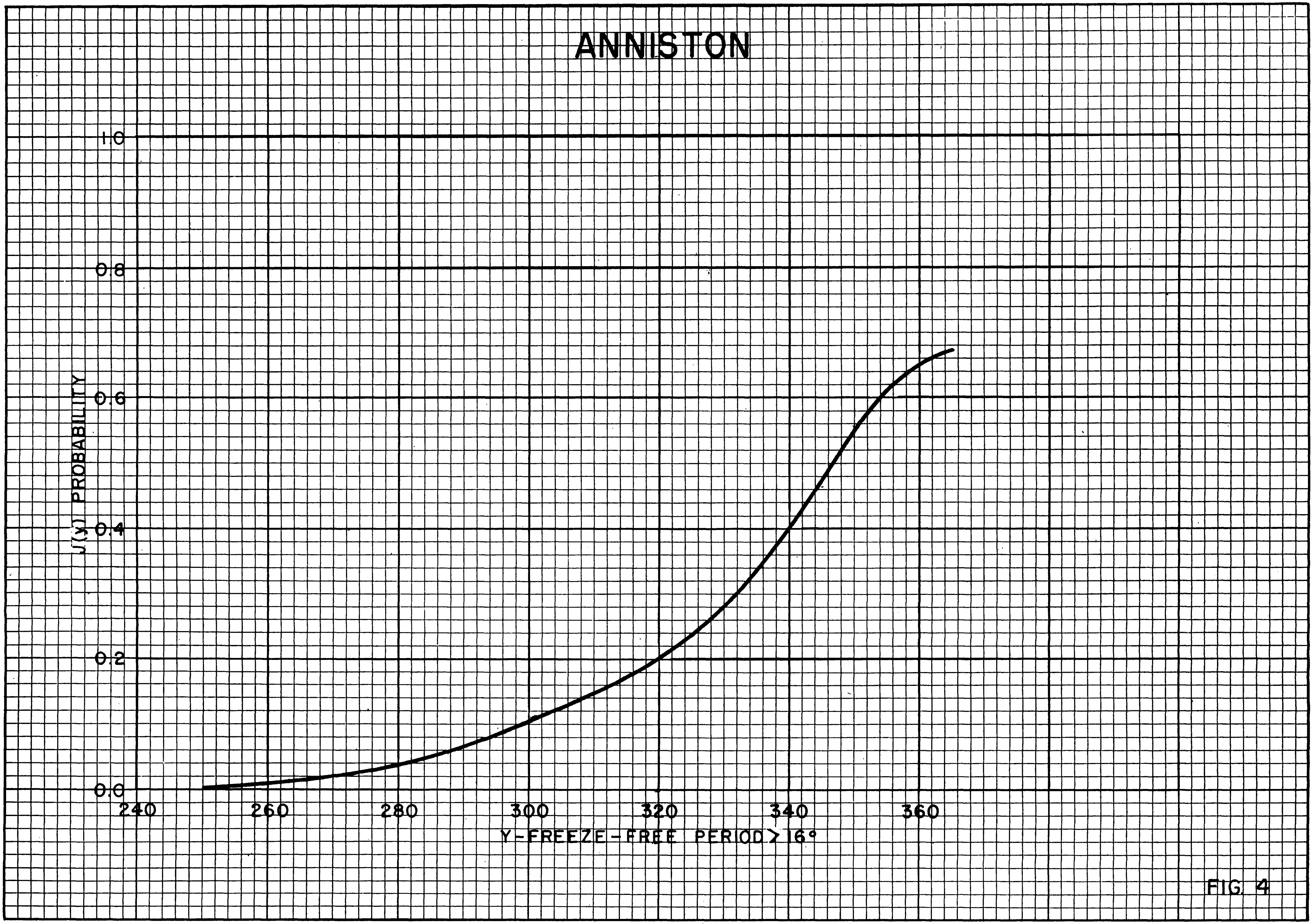

ANNISTON
J(y) PROBABILITY
0.0
0.2
0.4
0.6
0.8
1.0
240
260
280
300
320
340
360
Y-FREEZE-FREE PERIOD > 16°

FIG. 4